Enhancing Human Action Recognition with Region Proposals

Fahimeh Rezazadegan, Sareh Shirazi, Niko Sünderhauf, Michael Milford, Ben Upcroft
Australian Centre for Robotic Vision(ACRV), School of Electrical Engineering and Computer Science
Queensland University of Technology(QUT)
fahimeh.rezazadegan@qut.edu.au

Abstract

Deep convolutional network models have dominated recent work in human action recognition as well as image classification. However, these methods are often unduly influenced by the image background, learning and exploiting the presence of cues in typical computer vision datasets. For unbiased robotics applications, the degree of variation and novelty in action backgrounds is far greater than in computer vision datasets. To address this challenge, we propose an "action region proposal" method that, informed by optical flow, extracts image regions likely to contain actions for input into the network both during training and testing. In a range of experiments, we demonstrate that manually segmenting the background is not enough; but through active action region proposals during training and testing, state-of-the-art or better performance can be achieved on individual spatial and temporal video components. Finally, we show by focusing attention through action region proposals, we can further improve upon the existing state-of-the-art in spatio-temporally fused action recognition performance.

1 Introduction

Automatic recognition of human activities has been an active research area due to its potential application in a variety of domains. Human action recognition by mobile robots in real world scenarios is a challenging task that has been newly addressed in the robotics field. It could enhance the quality of service robots to identify their next required task or help robots report suspicious actions to keep the environment safe.

Recently deep learning has presented great performance for tasks such as object recognition [Krizhevsky et al., 2012; Razavian et al., 2014], face recognition [Taigman et al., 2014; Sun et al., 2013] and fine grained classification [Zhang et al., 2014], typically using static RGB images. For task recognition, an inherently motion-based field, researchers have been developing deep learning-based techniques which fuse conventional RGB images and optical flow information [Wang et al., 2015; Karpathy et al., 2014].

Recent two-stream Convolutional Neural Network (CNN) architectures [Simonyan and Zisserman, 2014; Ng et al., 2015; Donahue et al., 2015] have achieved state-of-the-art performance on action recognition benchmarks such as UCF101, UCF sport and HMDB datasets [Soomro et al., 2012; Kuehne et al., 2011], by fusing RGB and optical flow imagery.

Figure 1. Samples of some challenging video frames (UCF101 dataset) with contextually-informative backgrounds for recognizing human actions.

One of the challenges roboticists face in utilizing these systems on autonomous real world robots is that real world imagery is typically far more diverse and unbiased than computer vision datasets [Zhou et al., 2014]. This phenomenon is particularly apparent in action recognition, where traditional datasets tend to have contextually-informative backgrounds; an example being the standardized shot angles for sporting events (Figure 1). The research described in this paper is motivated by the need to develop generally deployable action recognition systems that work regardless of platform, context and background. Our overall approach is to develop a system which focuses on the regions in the image where actions are likely occurring, both at the training stage and during testing.

Our approach has three stages. The first utilizes optical flow to identify regions where action is occurring to provide action region proposals within the spatial imagery as well. In the second stage, the state-of-the-art network architecture is trained on these region proposals alone, rather than the full images. Finally, we fuse the learnt spatial and temporal features to produce a final classifier for action recognition. We conduct a range of experiments comparing the performance of our approach with the existing state-of-the-art systems, both with our region proposal system and with the control case of manual background removal. We show that the use of action region proposals results in matching or superior performance to the existing state-of-the-art, that manual removal of backgrounds during testing reduces the performance of full image-based state-of-the-art

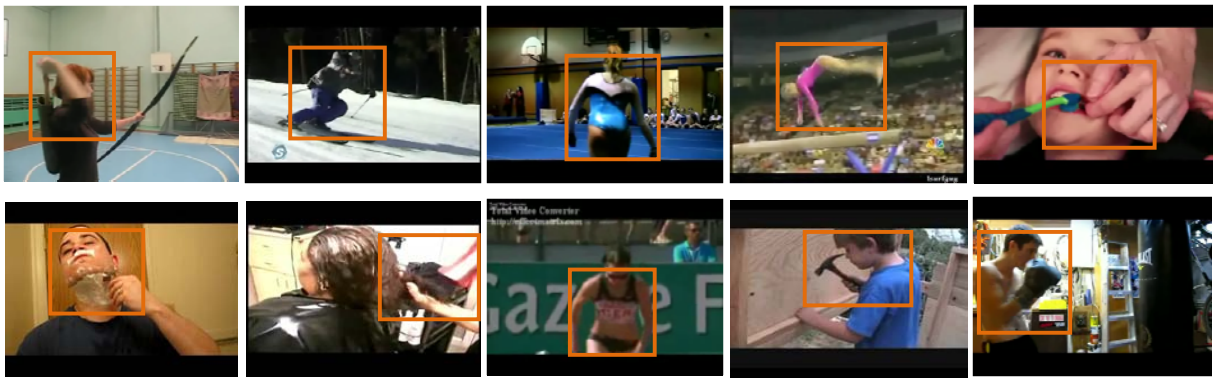

Figure 2. Samples of generated bounding boxes by our proposed method to detect action area before recognizing human actions.

techniques, and that fusing the spatial and temporal networks trained using action region proposals results in a new benchmark for action recognition.

The rest of paper is organized as follows. In Section 2, we review related work on action recognition. We then present an overview of our approach and describe our network training details in Section 3. Section 4 presents experimental results followed by conclusion in Section 5.

2 Related work

Recognizing human actions from videos is an important task for many applications such as video surveillance, human computer interaction and video content retrieval. In human action recognition, the common approach is to extract image features from the video and to issue a corresponding action class label. There has been a number of studies on human action recognition [Wang et al., 2013; Dollar et al., 2005; Laptev, 2005; Wang et al., 2014; Gkioxar et al., 2015] that can be categorized into two main groups of work: 1) hand crafted local features and bag of visual words representation. 2) deep learned feature descriptors. Both categories have demonstrated excellent results in recognition of human actions.

Recent work used shape-based features such as HOG [Dalal and Triggs, 2005], SIFT [Lowe, 2004] and motion dependent features such as optical flow, MBH [Dalal et al., 2006] with high order encodings (Bag of Words, Fischer vectors) and trained classifiers (e.g. SVM, decision forests) to predict actions.

Despite good performance in some cases, these hand-crafted descriptors are not optimized for visual representation and may lack discriminative capacity for action recognition.

Deep learning models are a class of machine learning algorithms that can learn a hierarchy of features by building high-level features from low-level ones. After impressive results of CNN on the task of image classification [Krizhevsky et al., 2012], researchers focused their effort mostly on proposing CNN models to solve action recognition problem as well [Baccouche et al., 2011; Ji et al., 2013; Wang et al., 2015; Du et al., 2015; Karpathy et al., 2014]. Recently proposed techniques such as Convolutional RBMs [Taylor et al., 2010], 3D CNNs [Ji et al, 2013], RNN [Du et al., 2015; Donahue et al., 2015], CNNs [Karpathy et al., 2014] and Two-Stream CNNs [Simonyan and Zisserman, 2014] have introduced valuable merits to this area.

Majority of research in action recognition considered optical flow as a local spatio-temporal feature, as well as appearance information to recognize human activities.

Ji et al. developed a 3D CNN model for action recognition instead of current 2D models [Ji et al, 2013]. In this work, features are extracted from both spatial and temporal dimensions by performing 3D convolutions, thereby capturing the motion information encoded in multiple adjacent frames. To attain the best performance, they also regularized outputs with high-level features and combined the predictions of a variety of different models [Kittler et al., 1998].

In the majority of recent work, temporal information has been employed to improve the result. In [Simonyan and Zisserman, 2014], a two stream CNN is proposed which has been the baseline of more recent studies [Donahue et al., 2015; Gkioxari and Malik, 2015; Wang et al., 2015]. In this paper, two spatial and temporal networks are combined. Spatial network mainly captures the discriminative appearance features for action understanding, while temporal network aims to learn the effective motion features. The proposed architecture, benefited from the late fusion. They also examined two different types of stacking techniques for its temporal network i.e. optical flow stacking and trajectory stacking. In other words, the horizontal and vertical flow channels ($d_t^{x,y}$) of L consecutive frames are stacked to form a total of 2L input channels which obtained the best result for L=10 or 20-channel optical flow images.

Traditional Recurrent Neural Networks (RNNs) can learn complex temporal dynamics by mapping input sequences to a sequence of hidden states, and hidden states to outputs. Recently, Donahue et al. proposed a long-term recurrent convolutional model [Donahue et al., 2015] that is applicable to visual time-series modeling. This work emphasizes that in visual tasks where spatial or temporal information has already been employed, long-term RNNs can provide significant improvement when ample training data are available to learn or refine the representation. Long Short Term memory (LSTM) architecture provides significant improvement for recognition on conventional video activity challenges. In this work, optical flow fields stacked in 3-channel images.

However, these deep models lack consideration of background effects on training images for the spatial network.

In this work, by selecting action region proposals as inputs to the network both during training and testing, we mainly focus on areas where actions are happening. It

not only improves the accuracy of the temporal network, but also prevents learning the cues from the background and provides more reliable result for robotics datasets.

3 Overview of the Approach

In this section, we describe our approach for enhancing human action recognition task. The summary of proposed method is also demonstrated in figure 3.

3.1 Selecting Action Region Proposals

Selecting action region proposals is a more challenging task compared to object proposals. This is mainly due to the fact that both appearance and motion cues are required to have a successful action region proposal, whereas object proposals are merely dependent on visual appearance information. Besides, considering the diversity of human actions, it is not straightforward to differentiate human actions from background and other dynamic motions [Yu and Yuan, 2015].

Our approach extracts areas of interest, action region proposals, at the frame level. Figure 2 shows generated region proposals of our method on samples of UCF101 dataset video frames. Motivated by EdgeBoxes technique [Lawrence and Dollár, 2014] that has been proved to perform well for object detection [Rezazadegan et al., 2015], we propose a strategy to choose the best action proposal. To this end, we slightly modify the EdgeBoxes method in order to detect appropriate action regions. We first extract video frames and then represent the motion using optical flow signals [Brox et al, 2004].

Applying EdgeBoxes on optical flow images, results in a large number of possible bounding boxes in an image which we must score efficiently for the specific task of action recognition. As a result, we score each bounding box based on the magnitude of optical flow signal within the box. In other words, we compute the score for each box using the normalized magnitude of the optical flow signal which can be considered as a heat map at the pixel level [Gkioxari and Malik, 2015]. The score function is:

$$OF(S) = \frac{1}{S}\sum_{i \in \Re} OF(i)$$

Where S is a bounding box. We discard S if $OF(S) \leq \delta$ that δ can be empirically attained. In our experiments, we choose δ=0.32 and set the EdgeBoxes parameters to the default values.

3.2 Training on Region Proposed Images

3.2.1 Network Architecture

Following the successful performance of AlexNet for image classification [Krizhevsky et al., 2012], other deep architectures such as VGGNet have been developed and demonstrated significant performance for large-scale image recognition [Simonyan and Zisserman, 2014].

This network has smaller convolutional kernel size (3×3), smaller convolutional strides (1×1), smaller pooling window (2×2) and deeper network architectures (16 and 19 layers).

Recently, VGGNet has been used in various studies for the task of action recognition.

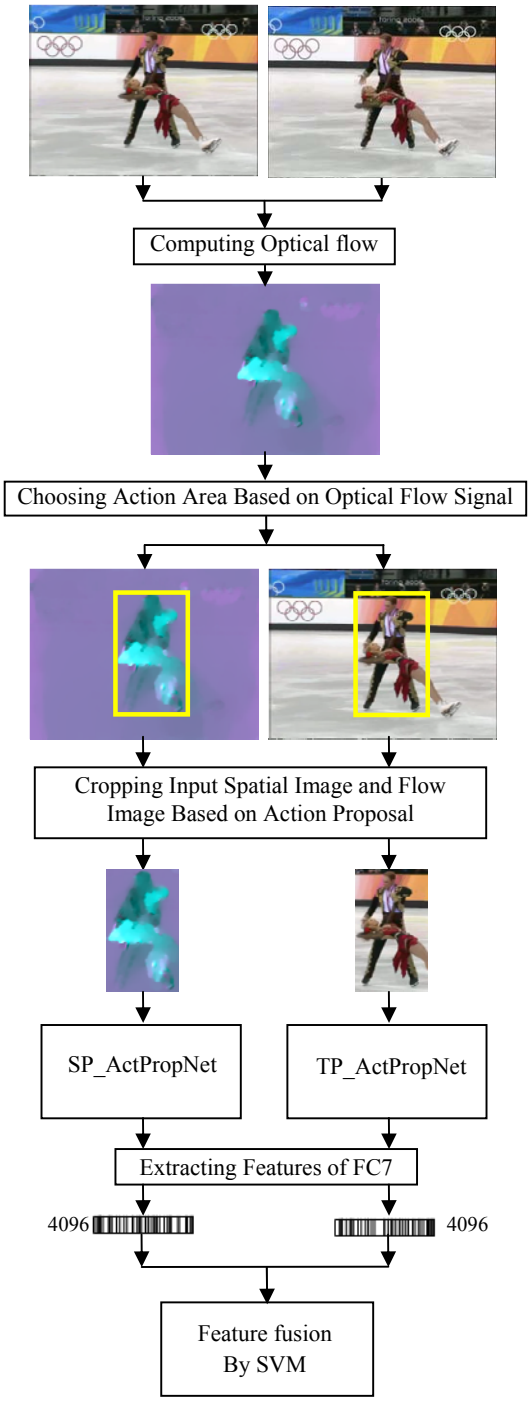

Figure 3. Summary of the proposed human action recognition approach based on action region proposals.

In this work, we utilize VGG-16 Layers architecture. This network architecture contains 13 convolutional layers and three fully connected layers. We train two individual networks, SP_ActPropNet and Tp_ActPropNet for region proposed images in spatial and temporal domains, respectively.

Same structure is employed for both spatial and temporal network (Figure 4). We used both region proposed images in spatial and optical flow domains in RGB format that is explained in section 3.2.2.

3.2.2 Network Training

In this section, we describe the process of training our deep action proposal network that is called ActPropNet. We train our model on UCF101 dataset [Soomro et al., 2012]. UCF101 is an action recognition dataset of realistic action videos, collected from YouTube, having 101 action categories http://crcv.ucf.edu/data/UCF101.php. It is a publicly available dataset containing 13320 video clips which organizers provided three splits into training and testing data, for evaluation. Each split has allocated almost 70% of video clips as training set and 30% as test set. We report the average of obtained accuracies on these three splits as the final accuracy in tables.

For each video clip, one frame is randomly selected and the horizontal (flow_x) and vertical optical flow signals (flow_y) are computed between two consecutive frames, for all selected frames. Mean subtraction is employed to reduce the effect of global motion between the frames.

As it is mentioned before, we employed an action region proposal to limit the background of images to where the action is occurring. Our action region proposal method is applied on optical flow images and then we use the generated bounding boxes for spatial images as well. Some recent methods extract the image regions by randomly cropping the full image [Simonyan and Zisserman, 2014; Donahue et al., 2015; Wang et al., 2015; Gkioxari and Malik, 2015], whereas we select image regions informed by optical flow which is more appropriate for motion-based tasks such as action recognition. Then region proposed images are resized to 224×224 before being fed to Caffe framework [Jia et al., 2014].

It should be emphasized that horizontal and vertical optical flow signals are saved in RGB format in which the third channel is formed by magnitude of optical flow signals and are linearly rescaled to a [0 255] range.

3.2.3 Implementation Details

The first network, SP_ActPropNet, is trained on region proposed spatial images which takes only the appearance clues of the scene. The second network, TP_ActPropNet, uses region proposed optical flow images generated by our action region proposal method. Both networks are trained with backpropagation, using Caffe framework [Jia et al., 2014]. We choose the ImageNet model as the initialization for both spatial and temporal network trainings. The learning rate is initially set to 0.001 and it is changed three times, during the training process. A momentum of 0.9 and a weight decay of 0.0005 is also used. We train the spatial network for 15K iterations because more iterations were unnecessary, due to the good initialization of the networks.

We implement temporal network training based on similar architecture, while learning rate of 0.005 is set initially and we change it five times until it reaches 40K iterations. The dropout rate for fully connected layers are also different from the spatial network. We used the dropout rate of 0.7 for layer FC6 and 0.9 for layer FC7.

3.2.4 Network Testing

At test time, we follow the same strategy proposed in [Simonyan and Zisserman, 2014] to have a fair comparison. We build our test data by extracting 25 spatial images and optical flow fields per video clip, to test the performance of SP_ActPropNet and TP_ActPropNet, respectively.

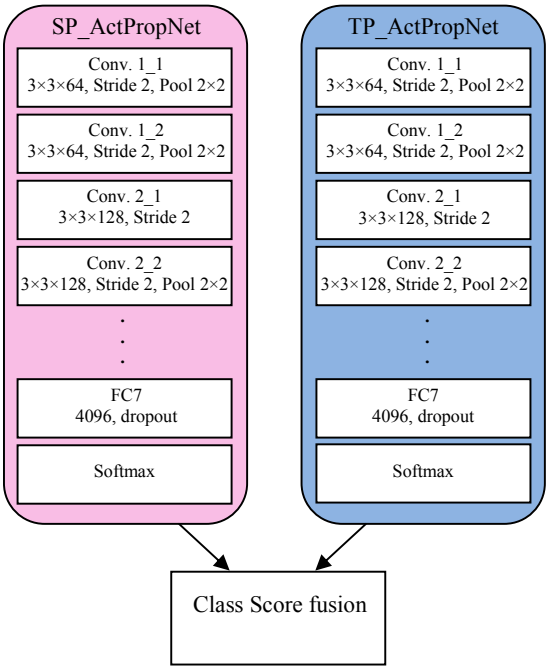

Figure 4. Our training network architecture for human action recognition.

3.3 Fusion of CNN Features

We use discriminative action classifiers on spatio-temporal features to make predictions for each region. The features are extracted from the final fully connected layer of the CNNs (FC7). We concatenate the CNN features from SP_ActPropNet and TP_ActPropNet, which is a 2×4096 dimensional descriptor, and then train a linear SVM as the final classifier. Figure 3 shows how spatial and motion cues are combined and fed into the SVM classifier.

4. Experiments and Results

We implemented the two-stream network proposed in [Simonyan et al., 2014] on UCF101 dataset. We achieved almost matching performance (72.1% for Spatial CNN and 72.6% for Temporal CNN in case of L=1).

Then we conducted another experiment using the trained spatial network to evaluate the performance in the control case of manual background removal.

Figure 5 shows how we replaced the background of some sample images in UCF101 dataset with white background. To do this experiment, we generated a small subset of test dataset (200 images), only including the foreground clues. After taking a test on trained spatial network (that we implemented following the instruction proposed by [Simonyan et al., 2014]) on the generated test set, performance dropped by approximately 8%.

Due to limited availability of robotics datasets, it would be beneficial to develop generally deployable action recognition systems by training the network on huge publicly available datasets and then refine the architecture based on the desired output task rather than training from scratch. Therefore, a reliable pre-trained model is required.

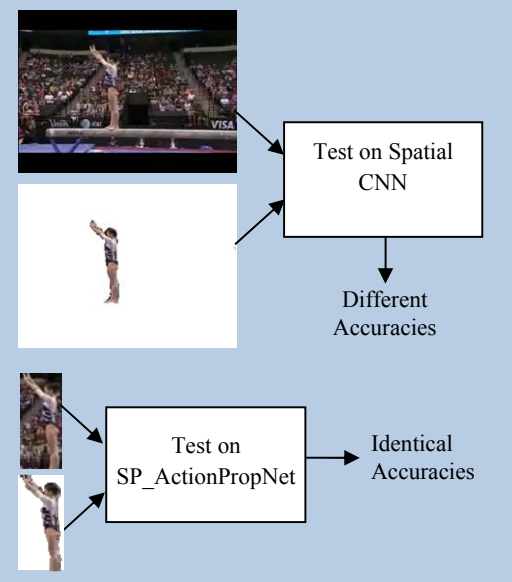

Figure 5. Manual background removal scenario.

To this end, we selected action regions of our training and test data as discussed in section 3.1 and performed the training process for both spatial and temporal networks, based on theses region proposed images instead of full images. Although the accuracy of spatial network was decreased slightly (2.6%) comparing with [Simonyan et al., 2014], but the accuracy of temporal network was highly increased (6.8%). However, lower accuracy for spatial stream proves the claim that their model uses the background cues to do the classification. After fusing the learnt spatial and temporal CNN features by SVM, we achieved matching accuracy (88.63%) to the state-of-the-art. Results have been demonstrated and compared with five state-of-the-art methods, in Table 1. Additionally, we carried out the same controlled experiment of manual background removal for testing SP_ActPropNet on the region proposed images extracted from the same newly generated test set. Interestingly, we observed that accuracy of SP_ActPropNet remained constant, which confirms our method is consistent regardless of the context and background. The results of background removal experiment for trained spatial network of [Simonyan et al., 2014] using full input images and SP_ActPropNet using region proposed input images have been shown in Table 2.

5 Conclusion

A reliable approach is proposed for action detection and recognition using convolutional neural networks based on appearance and motion cues. In a range of experiments, we demonstrated generic CNN models learned features from background clues as well as foreground information.

We sought a solution that was a generally applicable system regardless of the contextual information in the background. We developed an "action region proposal" method to automatically change the focus to the regions where the actions are likely happening. Through a number of experiments, we showed our temporal network outperformed the state-of-the-art using one optical flow field and our spatio-temporally fused action recognition performance matched or outperformed the state-of-the-art.

Training setting	Accuracy of Spatial Network on UCF101	Accuracy of Temporal Network on UCF101	Final Accuracy after fusion on UCF101
ActPropNet	70.1%	**80.7% (L=1)**	**88.63% (L=1)**
Two-stream CNN [1]	72.7%	73.9% (L=1) 81% (L=10)	N/A (L=1) 88% (L=10)
Single Frame [2]	69%	72.2%	79.04%
LRCN-fc6 [2]	71.12%	76.95%	82.95%
Two-stream +LSTM [3]	73.1%	N/A	88.6%
DeepNet [4]	65.4%		

Table 1. Performance comparison with the state-of-the-art deep networks on UCF101 dataset.

Model of Training	Accuracy on UCF101dataset	Accuracy on subset of 200 images with manually removed background
Implementation of Spatial CNN [1]	72.1%	64%
SP_ActPropNet	70.1%	**69.98%**

Table 2. Performance comparison of state-of-the-art work on UCF101 dataset with our model, before and after manual background removal task.

This work provides a more reliable trained model that has the capability of directly being transferred into real world robotics scenarios that experience diverse scenes. As future work, we are going to investigate how stacking multiple optical flow fields can improve the performance of the proposed approach.

Acknowledgements

This research was conducted by the Australian Research Council Centre of Excellence for Robotic Vision (project number CE140100016).

References

[Aggarwal and Ryoo, 2011] J. K. Aggarwal and M. S. Ryoo. Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3):1-43, 2011.

[Baccouche et al., 2011] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, A. Baskurt. Sequential deep learning for human action recognition. *HBU Springer*, pages 29–39, 2011.

[Brox et al., 2004] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *Proceedings of the*

1 Simonyan and Zisserman, NIPS, 2014
2 Donahue et al., CVPR, 2015
3 Ng et al., CVPR, 2015
4 Karpathy et al., CVPR, 2014

European Conference on Computer vision (ECCV), pages 25–36, 2004.

[Dalal and Triggs, 2005] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2005.

[Dalal et al., 2006] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In *Proceedings of the European Conference on Computer vision (ECCV)*, 2006.

[Deng et al., 2012] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Competition2012 (ILSVRC2012).

[Doll´ar et al., 2005] P. Doll´ar, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In *VS-PETS*, 2005.

[Du et al., 2015] Y. Du, W. Wang, L. Wang. Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2015.

[Gkioxari and Malik, 2015] G. Gkioxari and J. Malik. Finding action tubes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[Gkioxari et al., 2015] G. Gkioxari, R. Girshick, and J. Malik. Contextual action recognition with r* cnn. arXiv:1505.01197, 2015.

[Jhuang et al., 2013] H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. Black. Towards understanding action recognition. In *Proceedings of the IEEE International Conference On Computer Vision (ICCV)*, 2013.

[Ji et al., 2013] S. Ji, W. Xu, M. Yang, and K. Yu. 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, January 2013.

[Jia et al., 2014] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *ACM MM*, 2014.

[Karpathy et al., 2014] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2014.

[Kittler et al., 1998] J. Kittler, M. Hatef, R.P. Duin, and J. Matas. On Combining Classifiers. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(3):226-239, Mar. 1998.

[Krizhevsky et al., 2012] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, pages 1106–1114, 2012.

[Kuehne et al., 2011] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. HMDB: A large video database for human motion recognition. In *Proceedings of the IEEE International Conference On Computer Vision (ICCV)*, 2011.

[Laptev, 2005] Laptev. On space-time interest points. *IJCV*, 64(2/3):107-123, 2005.

[Lawrence and Dollár, 2014] C. Lawrence Zitnick and Piotr Dollár. Edge boxes: Locating object proposals from edges. In *Proceedings of the European Conference on Computer vision (ECCV)*, 2014.

[Lowe, 2004] DG Lowe. Distinctive image features from scale-invariant key points. In IJCV, 2004.

[Ng et al., 2015] J. Y.-H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici. Beyond short snippets: Deep networks for video classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pages 4694–4702, 2015.

[Razavian et al., 2014] A. Razavian, H. Azizpour, J. Sullivan, S. Carlsson. CNN Features off-the-shelf: an Astounding Baseline for Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* 2014.

[Rezazadegan et al., 2015] F. Rezazadegan, S. Shirazi, M. Milford, B. Upcroft. Evaluation of object detection proposal under condition variations. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* 2015.

[Simonyan and Zisserman, 2014] K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, pages 1-8, 2014.

[Simonyan and Zisserman, 2014] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR, abs/1409.1556*, 2014.

[Soomro et al., 2012] K. Soomro, A. R. Zamir, and M. Shah. UCF101: A dataset of 101 human actions classes from videos in the wild. *CoRR, abs/1212.0402*, 2012.

[Sun et al., 2014] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[Taigman et al., 2014] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[Taylor et al., 2010] G.W. Taylor, R. Fergus, Y. LeCun, and C. Bregler. Convolutional learning of spatio-temporal features. In *Proceedings of the European Conference on Computer vision*, 2010.

[Wang et al., 2013] H. Wang, A. Kläser, C. Schmid, C.L. Liu. Dense Trajectories and Motion Boundary Descriptors for Action Recognition. 2013, *Int J Comput Vis*, 103:60–79, 2013.

[Wang et al., 2014] L. Wang, Y. Qiao, and X. Tang. Latent hierarchical model of temporal structure for complex activity classification. *TIP*, 23(2), 2014.

[Wang et al., 2015] L. Wang, Y. Qiao, X. Tang. Action Recognition with Trajectory-Pooled Deep-Convolutional Descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2015.

[Yu and Yuan, 2015] G. Yu and J. Yuan. Fast action proposals for human action detection and search. In *Proceedings of the IEEE Conference on Computer* Vision and Pattern Recognition (CVPR), 2015.

[Zhang et al., 2014] N. Zhang, J. Donahue, R. Girshick, and T. Darrell. Partbased r-cnns for fine-grained category detection. In *Computer Vision–ECCV*, pages 834–849. *Springer*, 2014.

[Zhou et al., 2014] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Object Detectors Emerge in Deep Scenes CNNs, *arXiv:1412.6856 [cs.CV].*